



# Churn Analysis

---

Team : Julien Bouhelier

# Sommaire



■ Business context

■ ML analyses

■ Statistics

■ Business  
recommendations

■ Data préparation



# Business Context

## Churn



Le churn est un kpi essentiel pour comprendre la satisfaction et la fidélité. En suivant le taux d'attrition, vous pourrez identifier les domaines où les clients ne sont pas satisfaits et apporter des changements pour améliorer l'expérience client.

## CAC



Le coût d'acquisition du client (CAC) est le coût total du marketing et des ventes pour acquérir un nouveau client. Dans votre cas (Banque), le CAC est élevé, il est donc essentiel de conserver vos clients et de réduire le churn.

## Compréhension

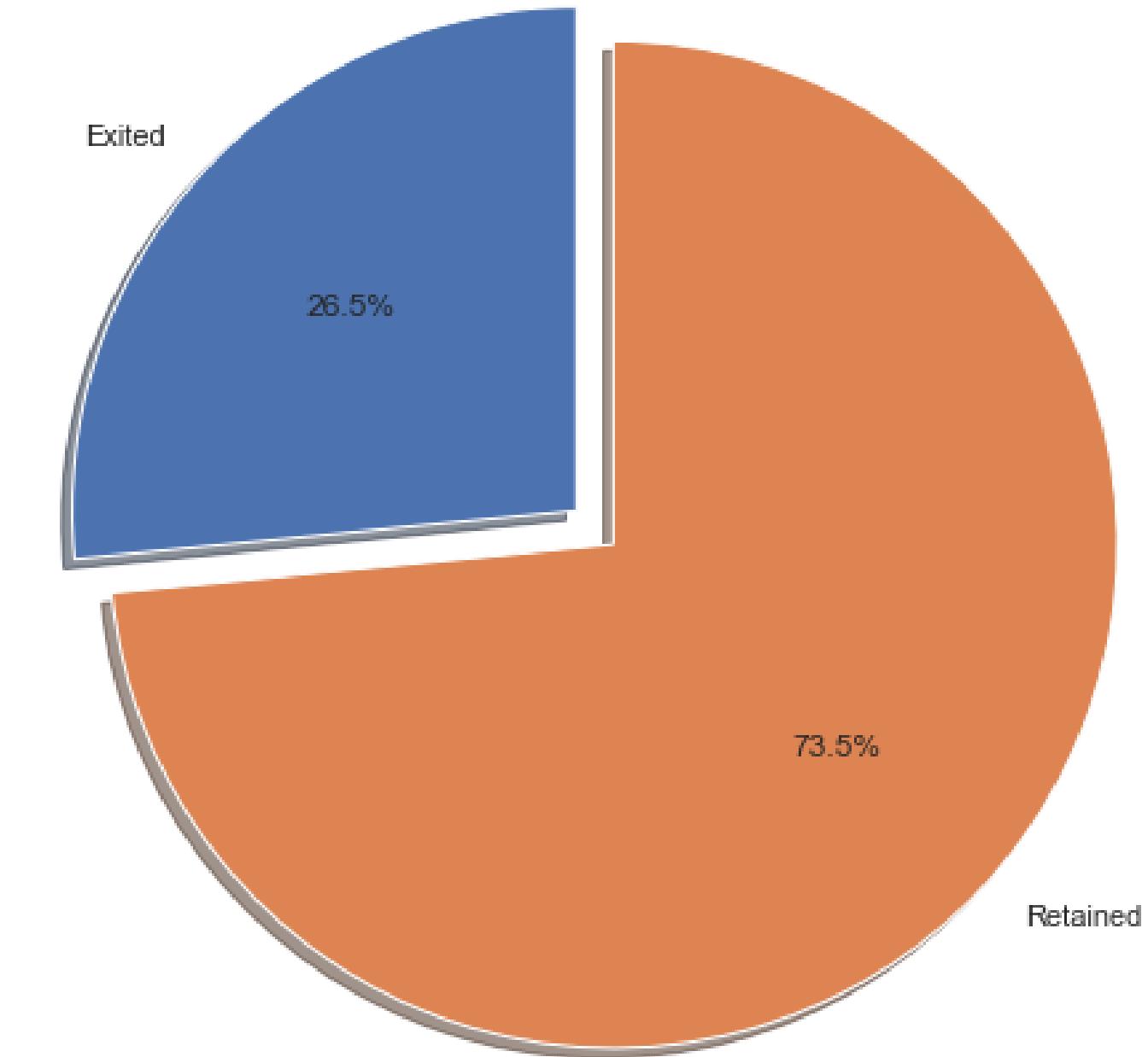


Grâce à cette analyse, vous connaîtrez mieux vos clients. Cette compréhension est clé pour améliorer le service, et donc acquérir de nouveaux clients grâce à des recommandations positives et du bouche à oreille par exemple.

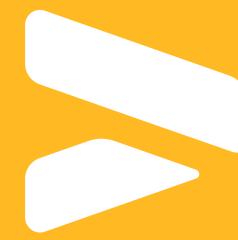


# Data

Proportion of customer churned and retained

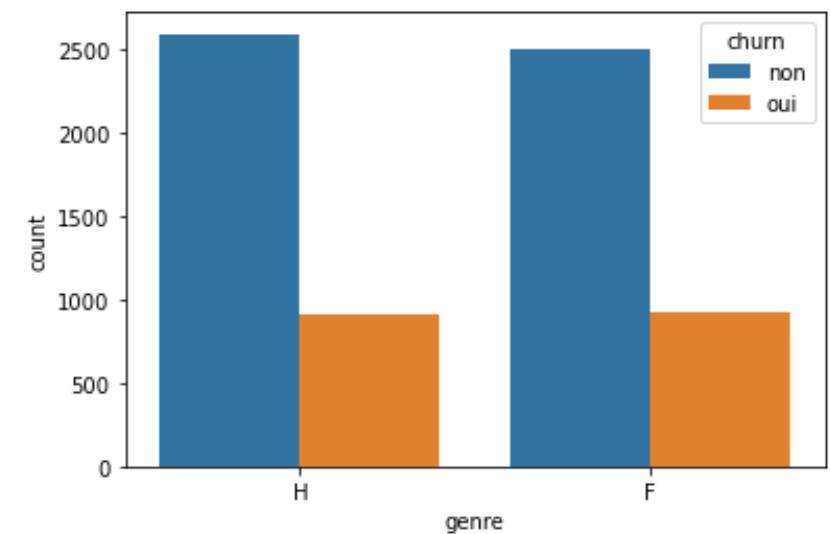
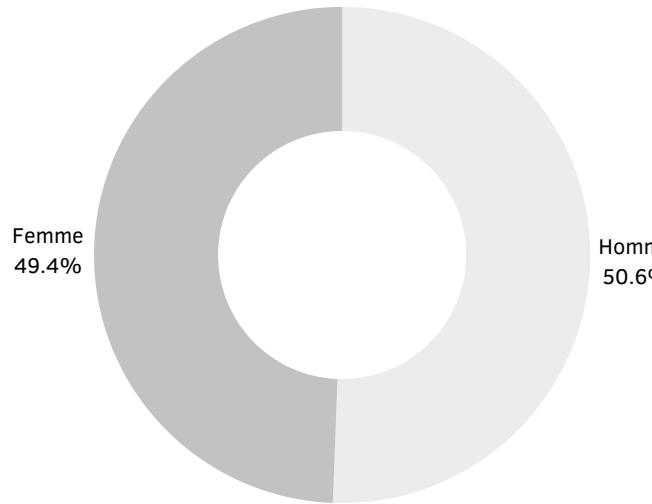


Remarque : oversampling

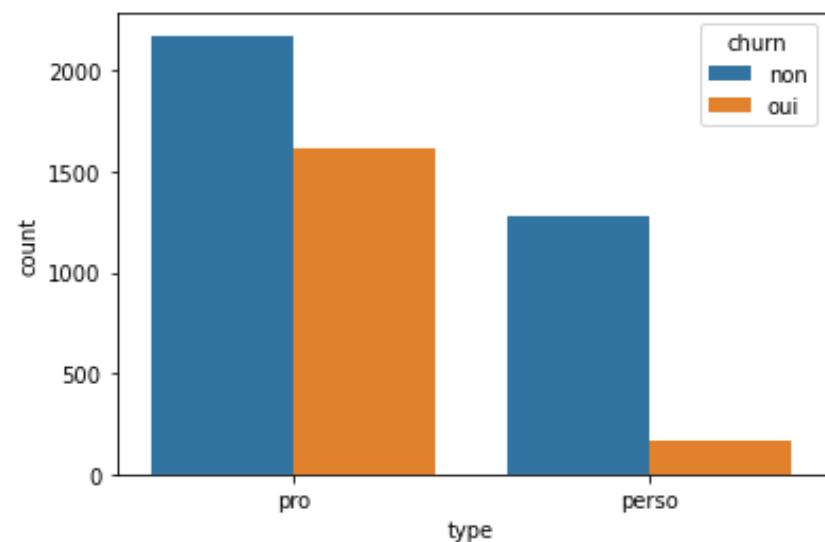
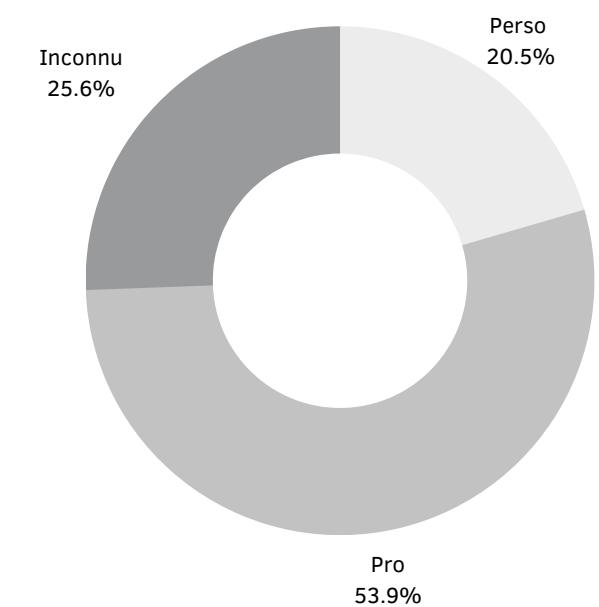


# Clientèle

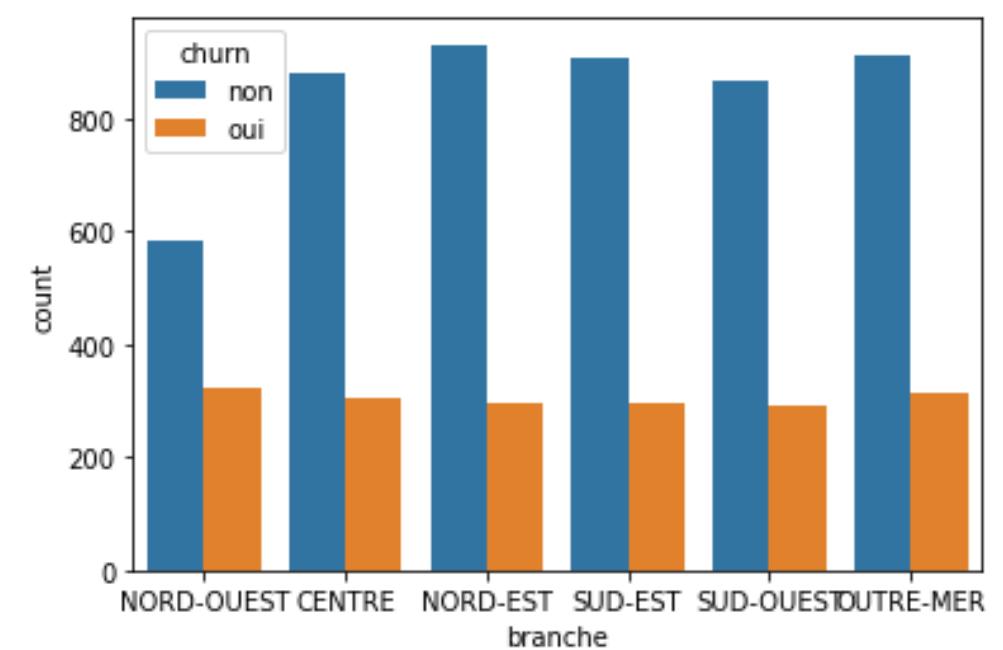
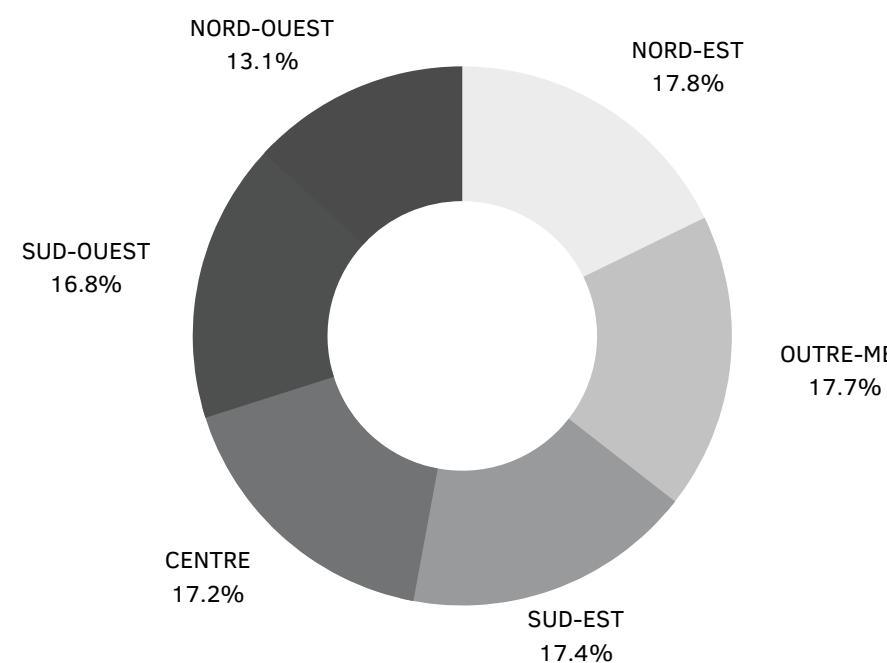
## Genre



## Contrat



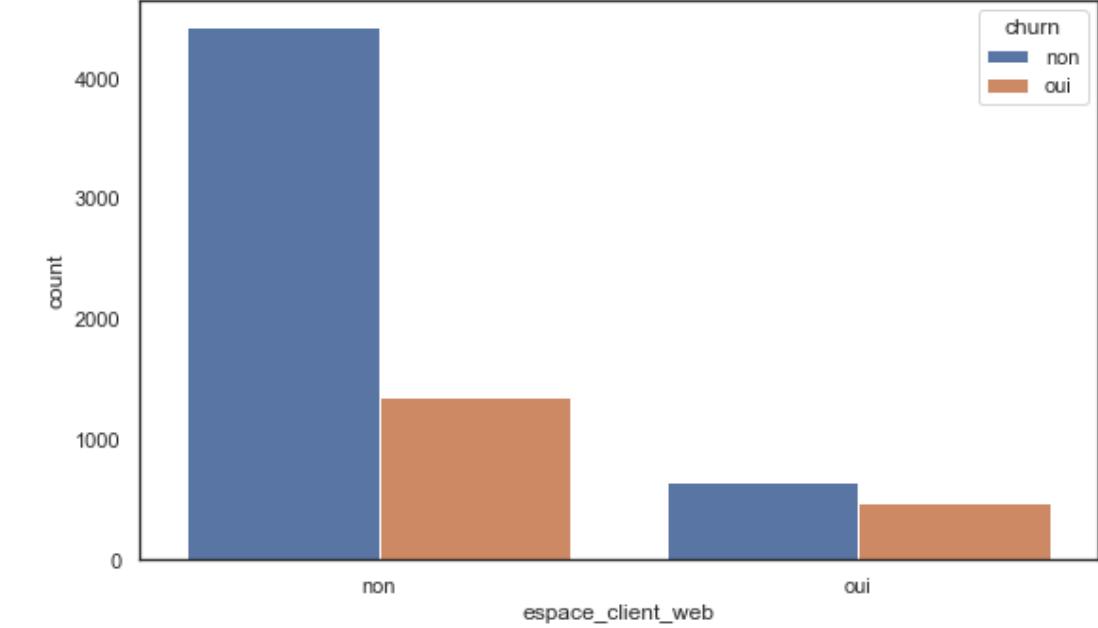
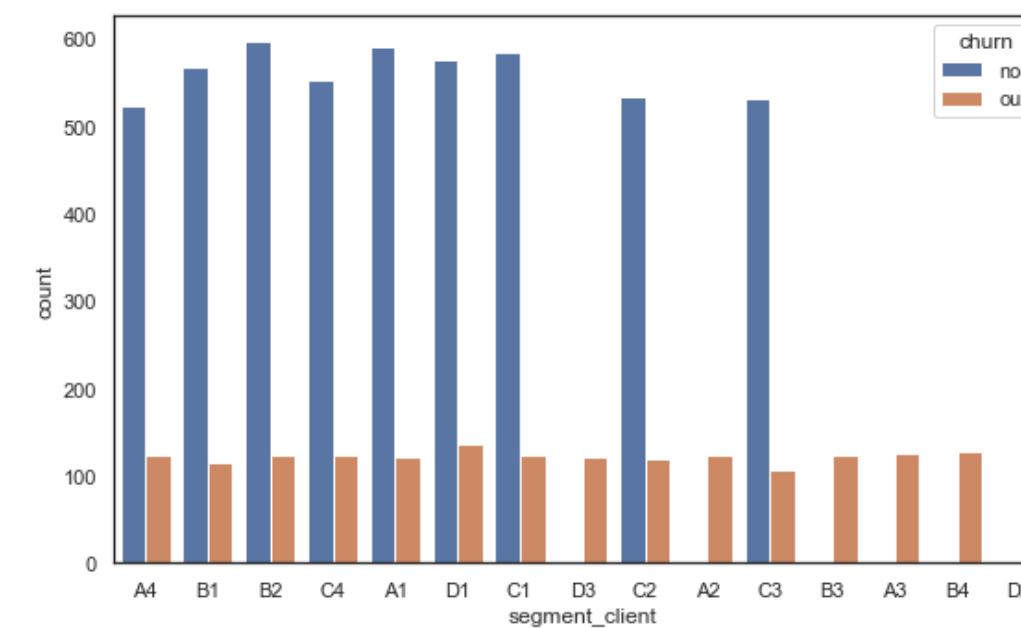
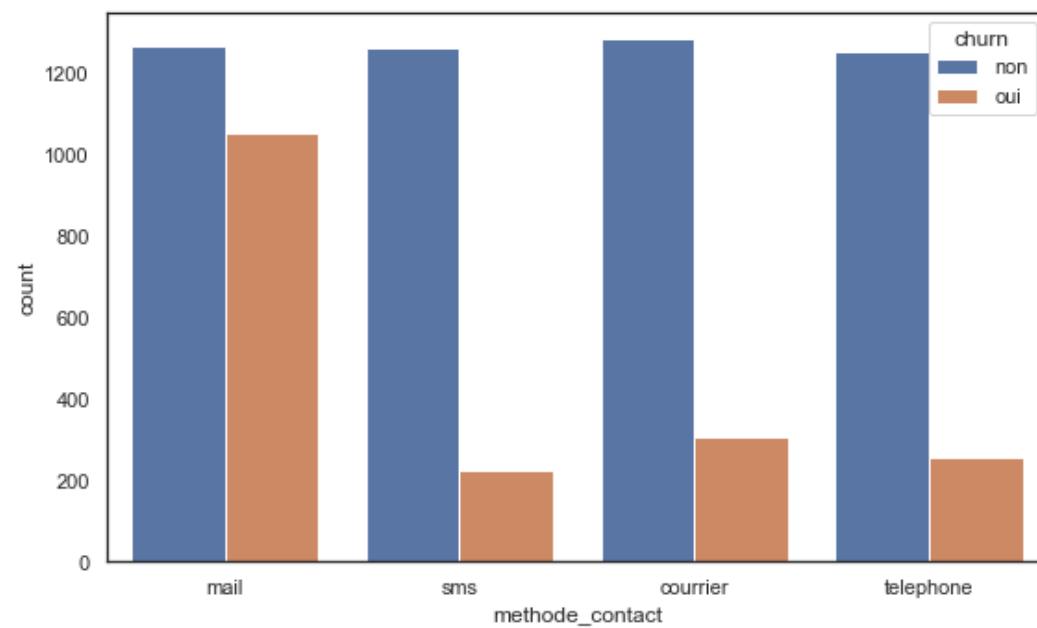
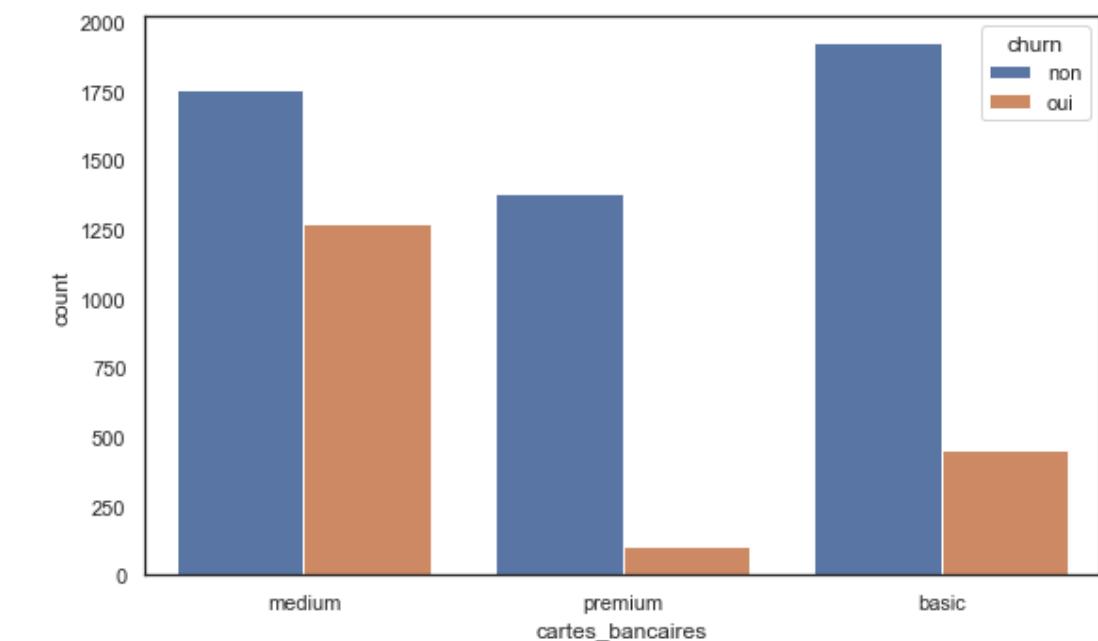
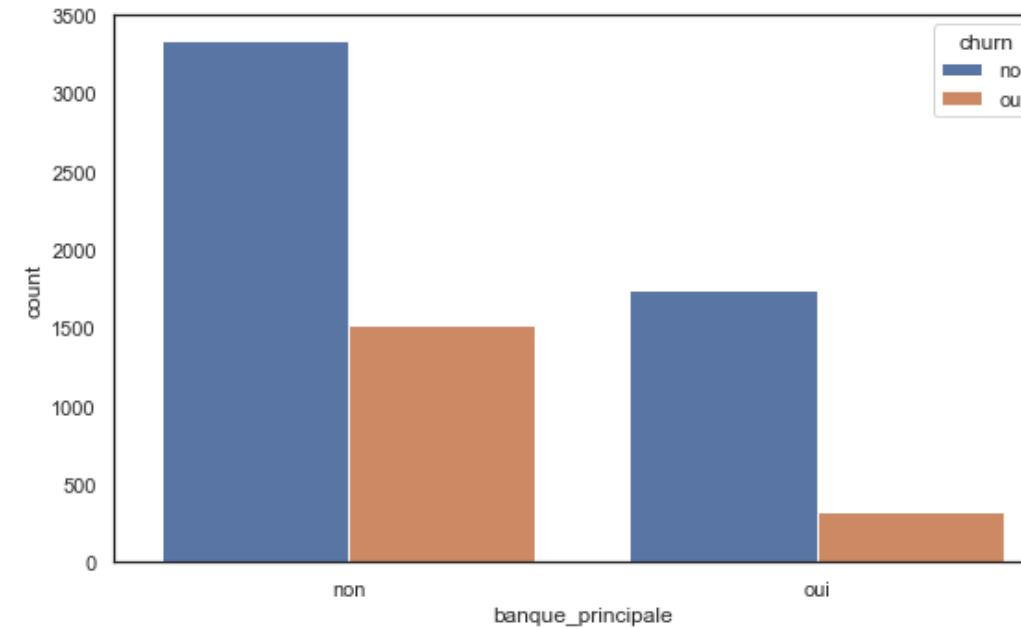
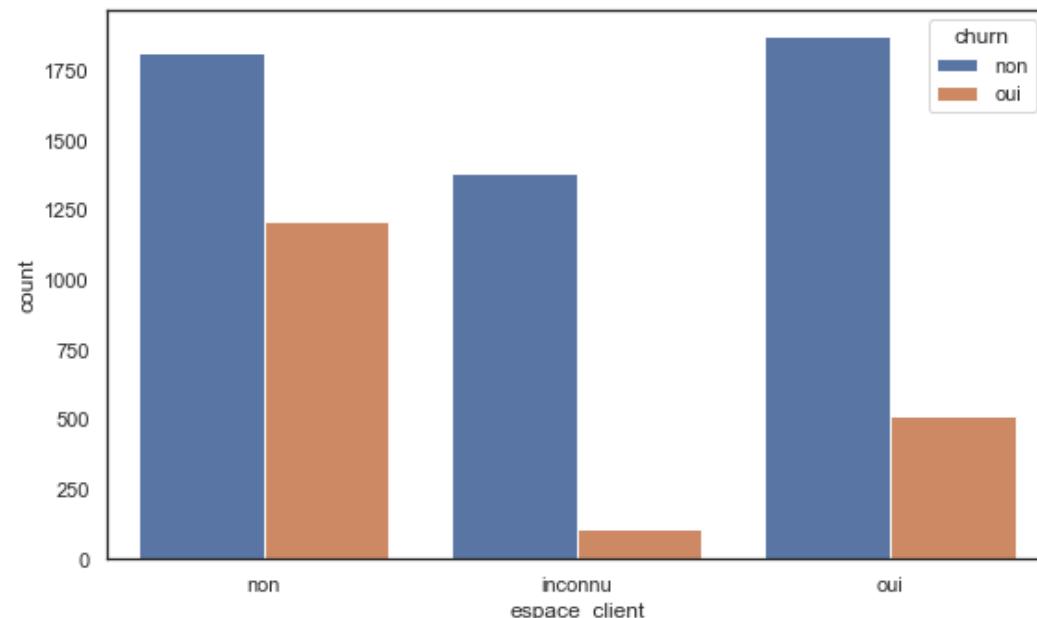
## Branche

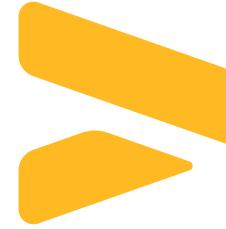




# Analyse : Influence sur le churn

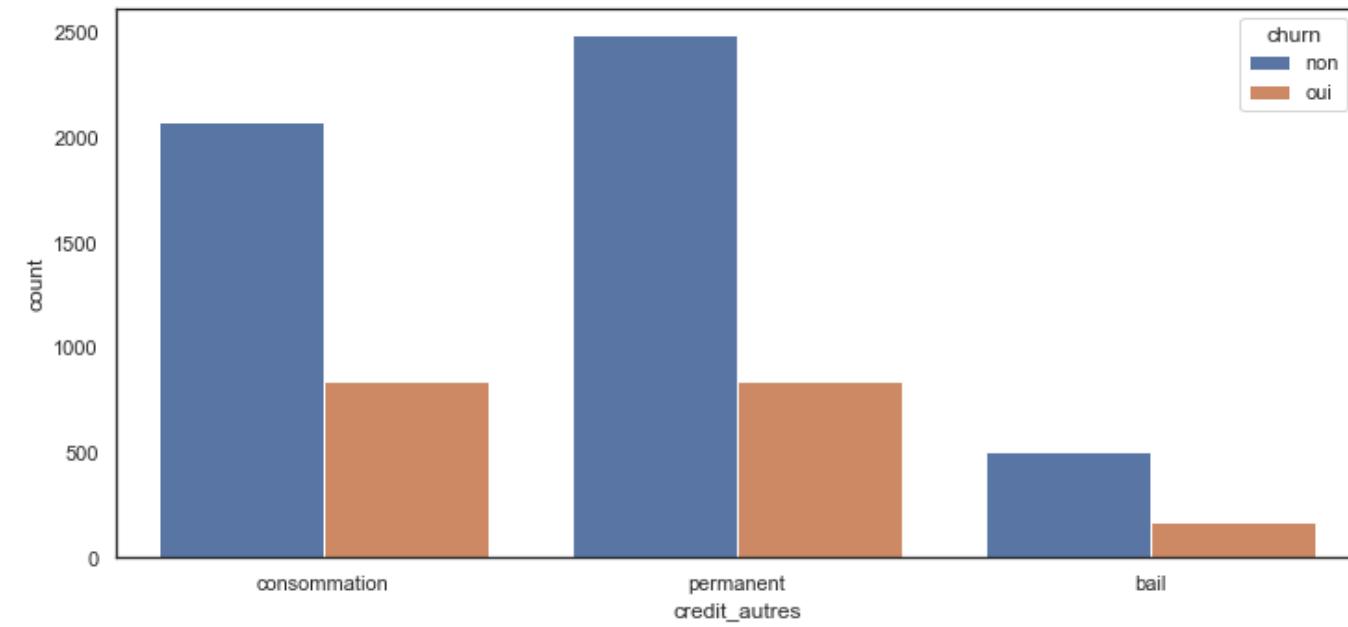
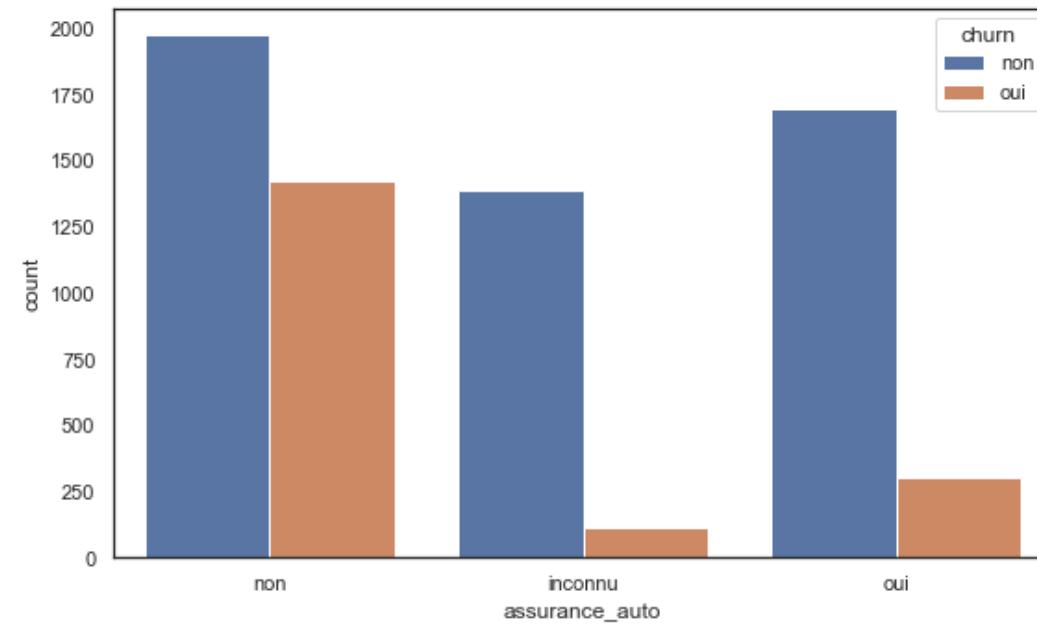
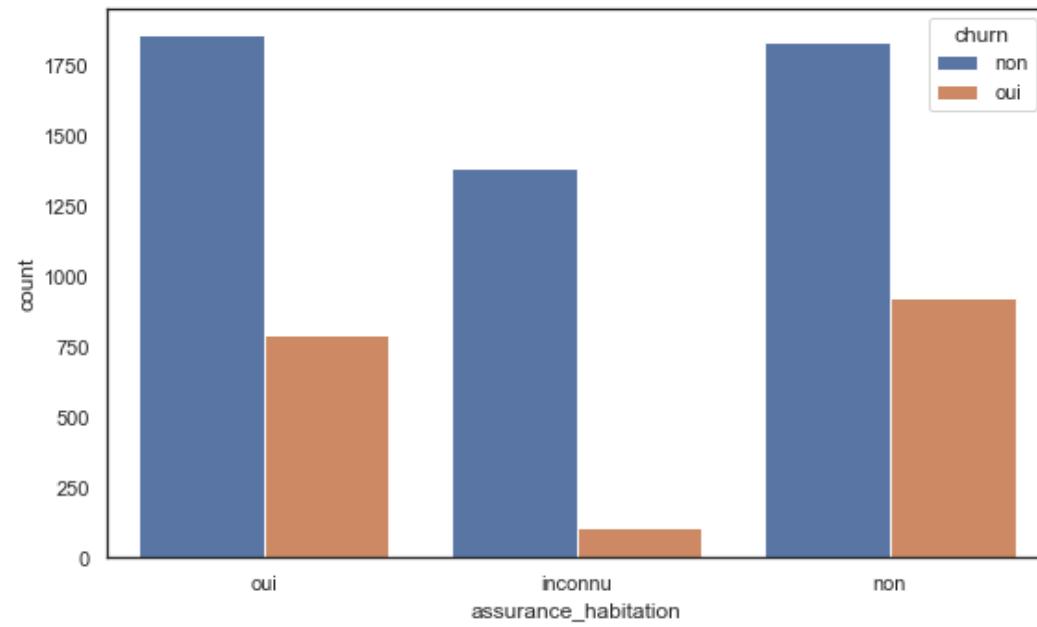
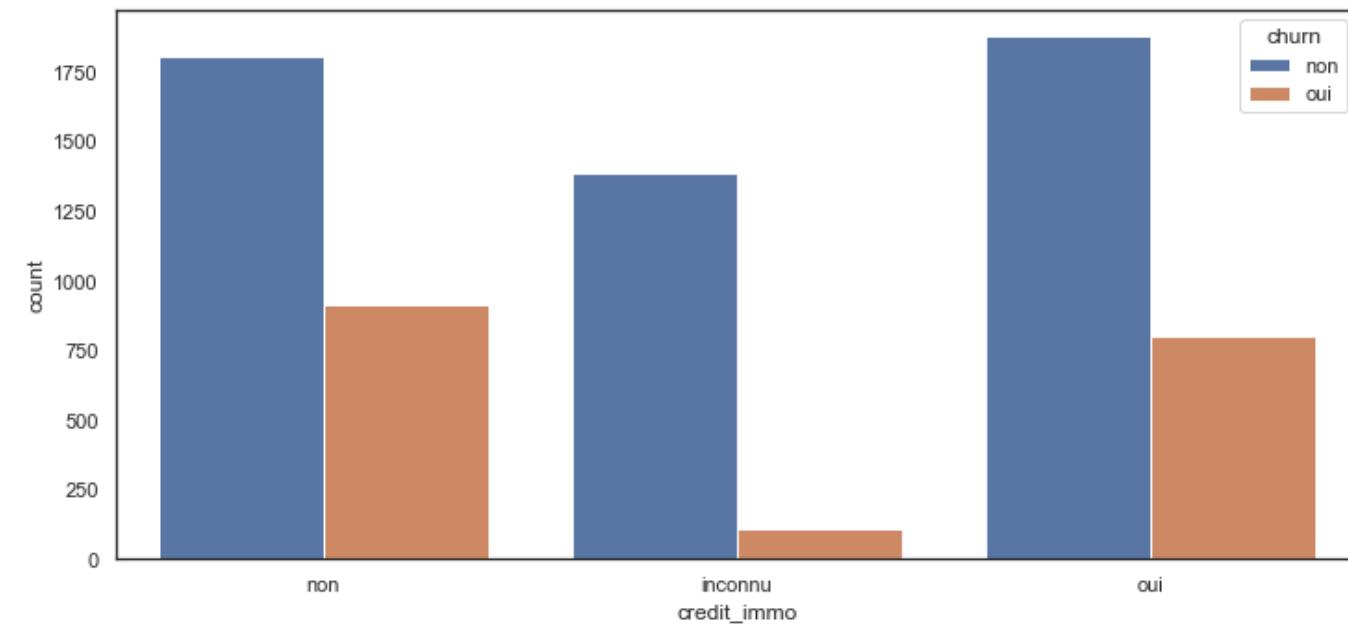
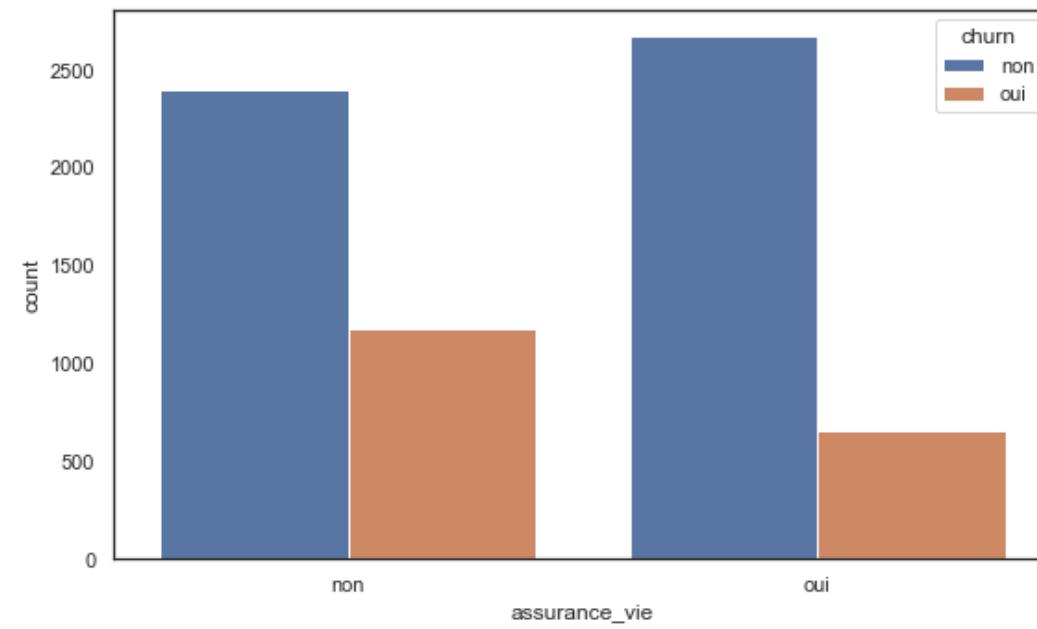
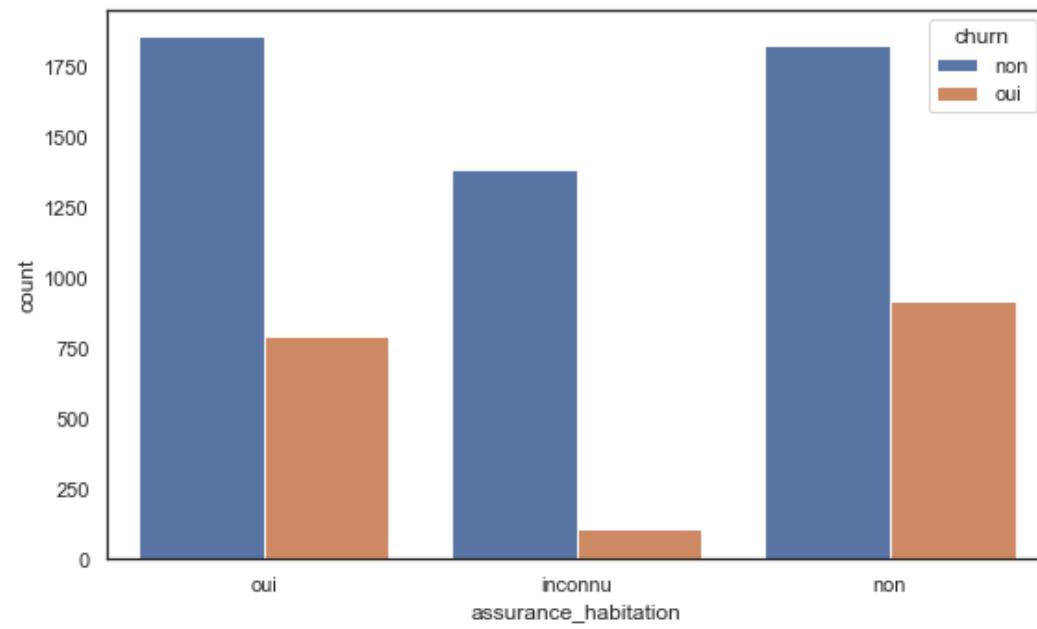
## Categorical Variables : General

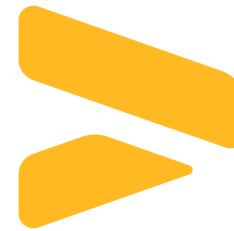




# Analyse : Influence sur le churn

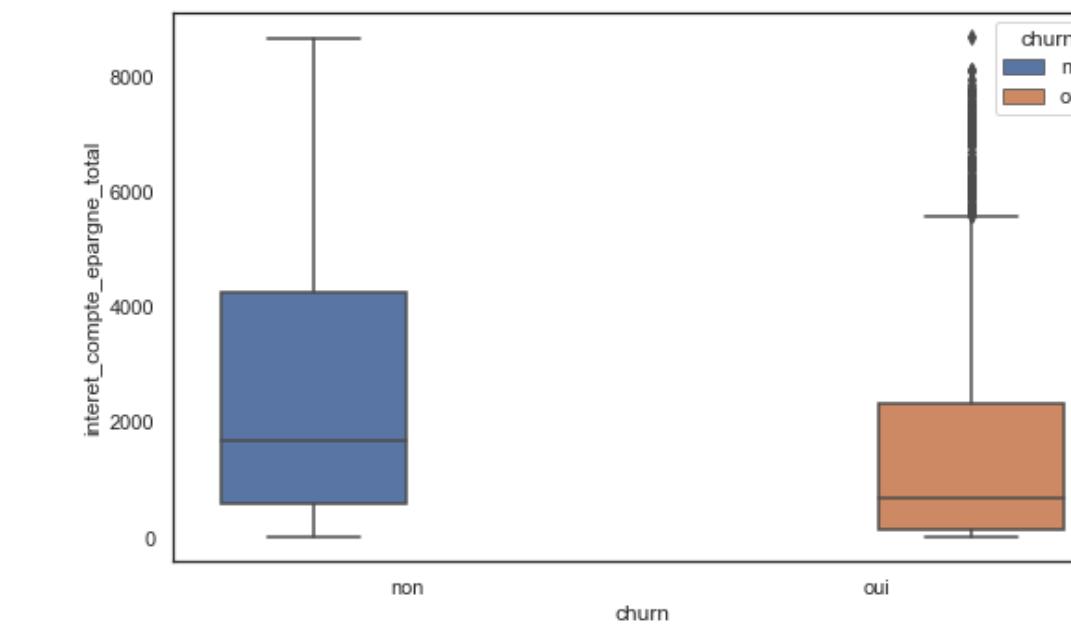
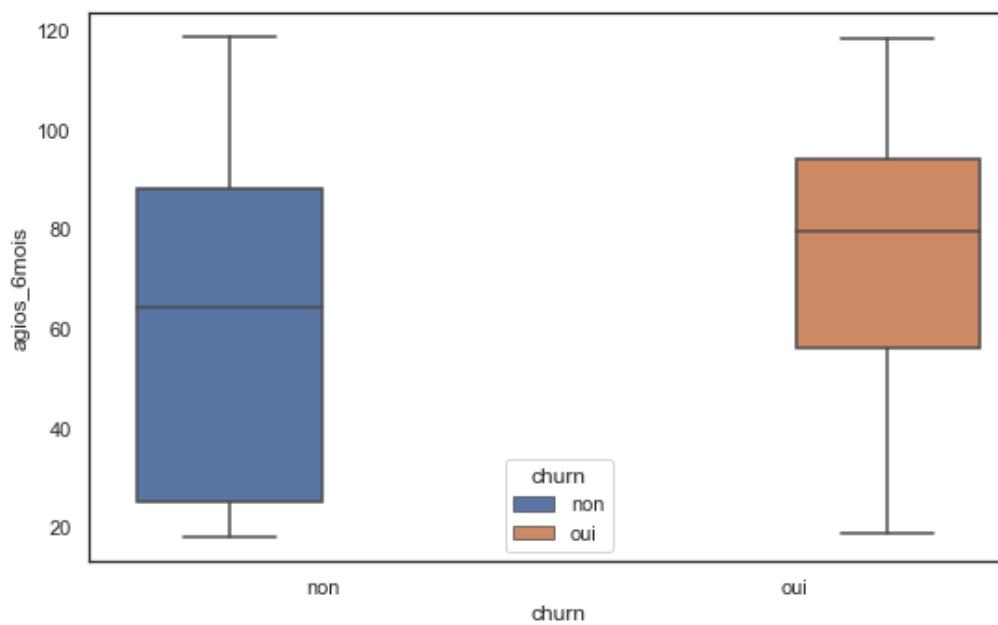
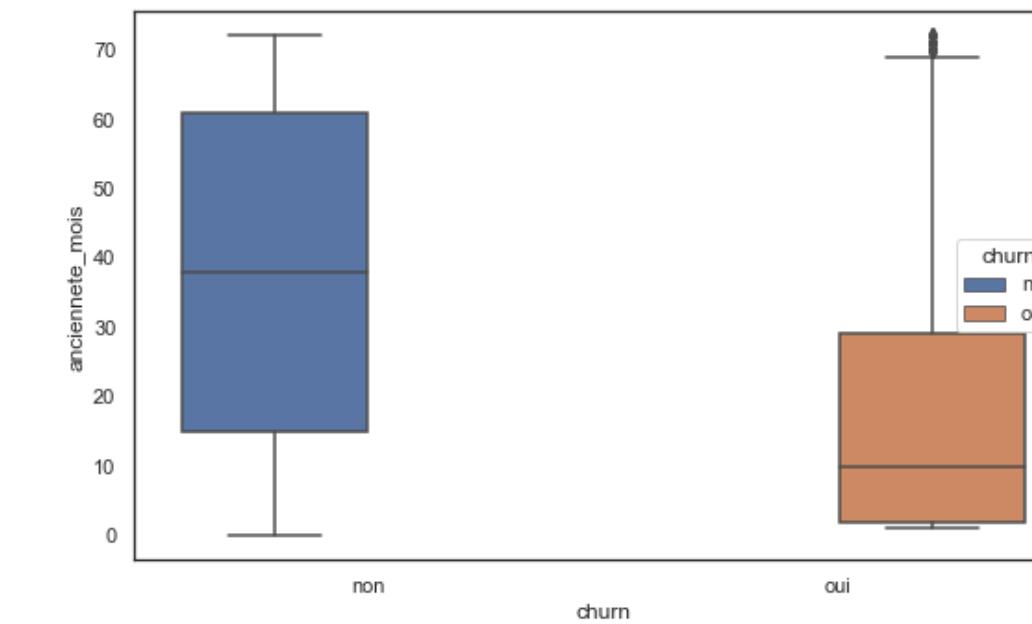
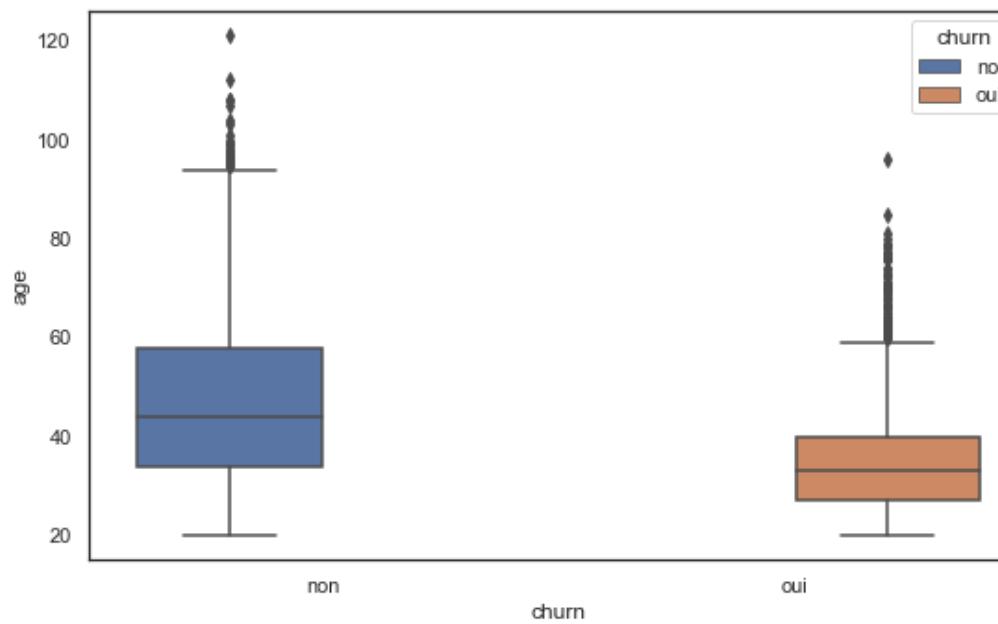
## Categorical Variables : Assurances et crédits

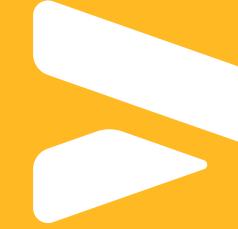




# Analyse : Influence sur le churn

Variables continues :





# Analyses : Variables Retenues

1

## VARIABLES SANS INFLUENCE

On ne prendra pas pour construire notre modèle le genre, la branche and et les credits car ils n'affectent pas le churn

2

## VARIABLES AVEC TROP DE D'INCONNUS

On ne vas pas garder le type (Perso ou pro car 25% des lectures manquantes)

Problème avec les cartes bancaires : Premium systématiquement manque d'infos sur beaucoup de variables :

Index	ale	espace client web	compte epargne	compte titres	assurance vie	segment client	methode contact	cartes bancaires	espace client	compte courant	PFA	assurance habitation	assurance auto
15	non	oui	non	non	D1	courrier	premium	inconnu	inconnu	inconnu	inconnu	inconnu	inconnu
16	non	oui	oui	non	B2	sms	basic	non	non	non	non	non	non
18	non	oui	oui	oui	B2	telephone	basic	non	oui	oui	oui	oui	oui
19	non	oui	non	non	C2	sms	basic	non	non	non	oui	oui	non
20	non	oui	oui	oui	B2	mail	medium	non	non	non	non	non	oui
23	non	oui	non	oui	C4	courrier	premium	inconnu	inconnu	inconnu	inconnu	inconnu	inconnu
24	non	oui	non	oui	A4	sms	premium	inconnu	inconnu	inconnu	inconnu	inconnu	inconnu
25	oui	oui	oui	oui	B1	telephone	medium	oui	non	oui	oui	oui	non
26	non	non	oui	oui	C1	sms	basic	non	non	non	non	non	oui

3

## VARIABLES AVEC VALEURS MANQUANTES

Nous prenons la décision de ne pas garder les valeur les valeurs anonymes car trop de valeurs manquantes (Quand on les supprimés on se retrouve avec un dataset de 2000 lectures seulement)

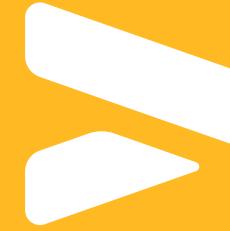
Remarque : On aurait pu remplacer par la moyenne (Environ 150 valeurs manquantes par variables mais nous avons fait des modèles avec, cela n'améliore pas le résultat.

De plus, pas interprétables.

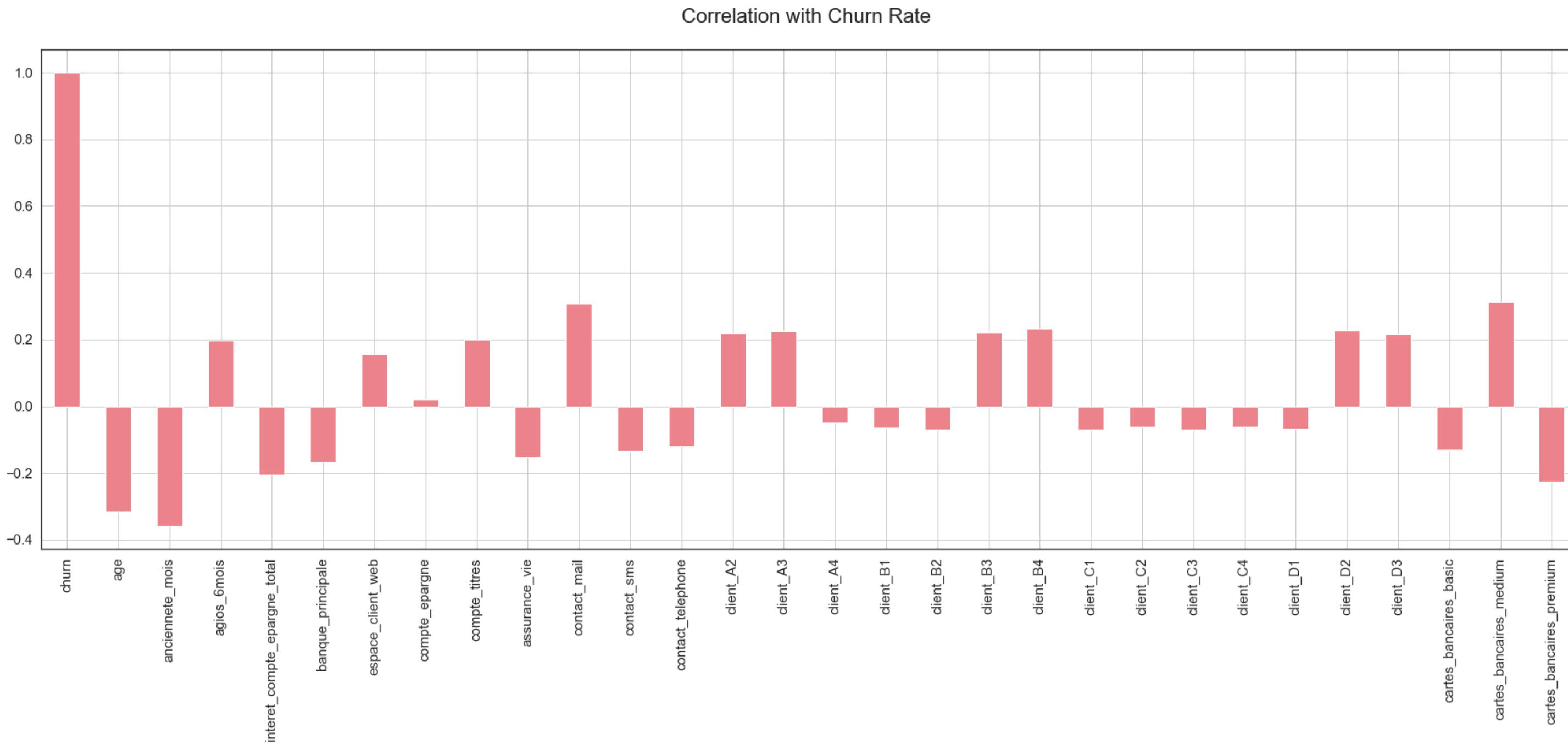


# Data Prep for Modelling





# Correlation : Variables



## INTERPRETATION

Valeur positive :

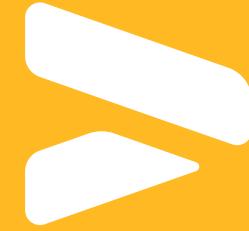
- Augmente le churn

Valeur négative :

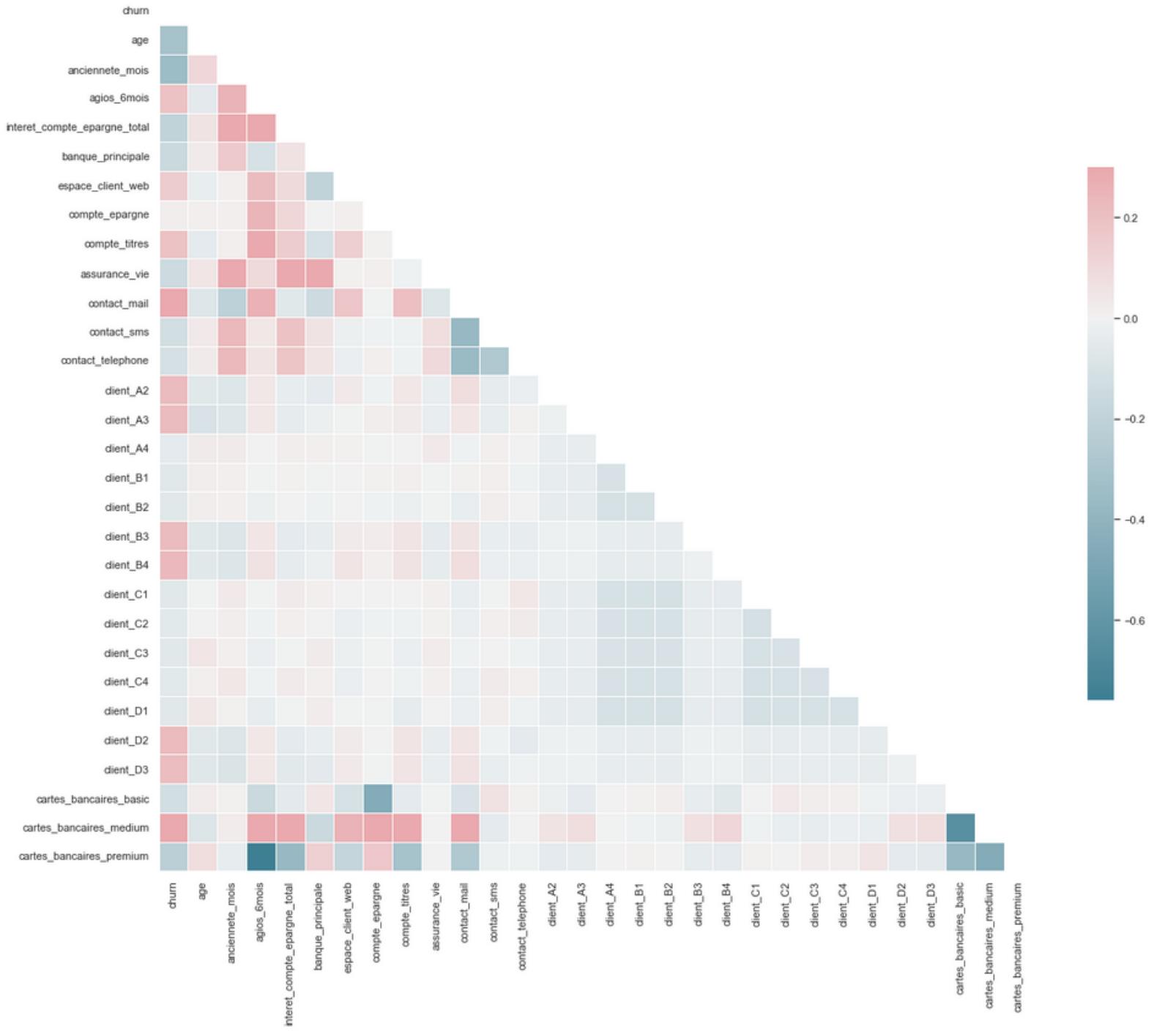
- Diminue le churn

## DUMMIES VARIABLES

Ici on a transformé les variables "oui" ou "non en 1 ou 0 binaires



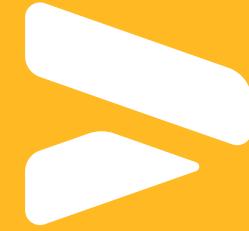
# Colinéarité entre les variables



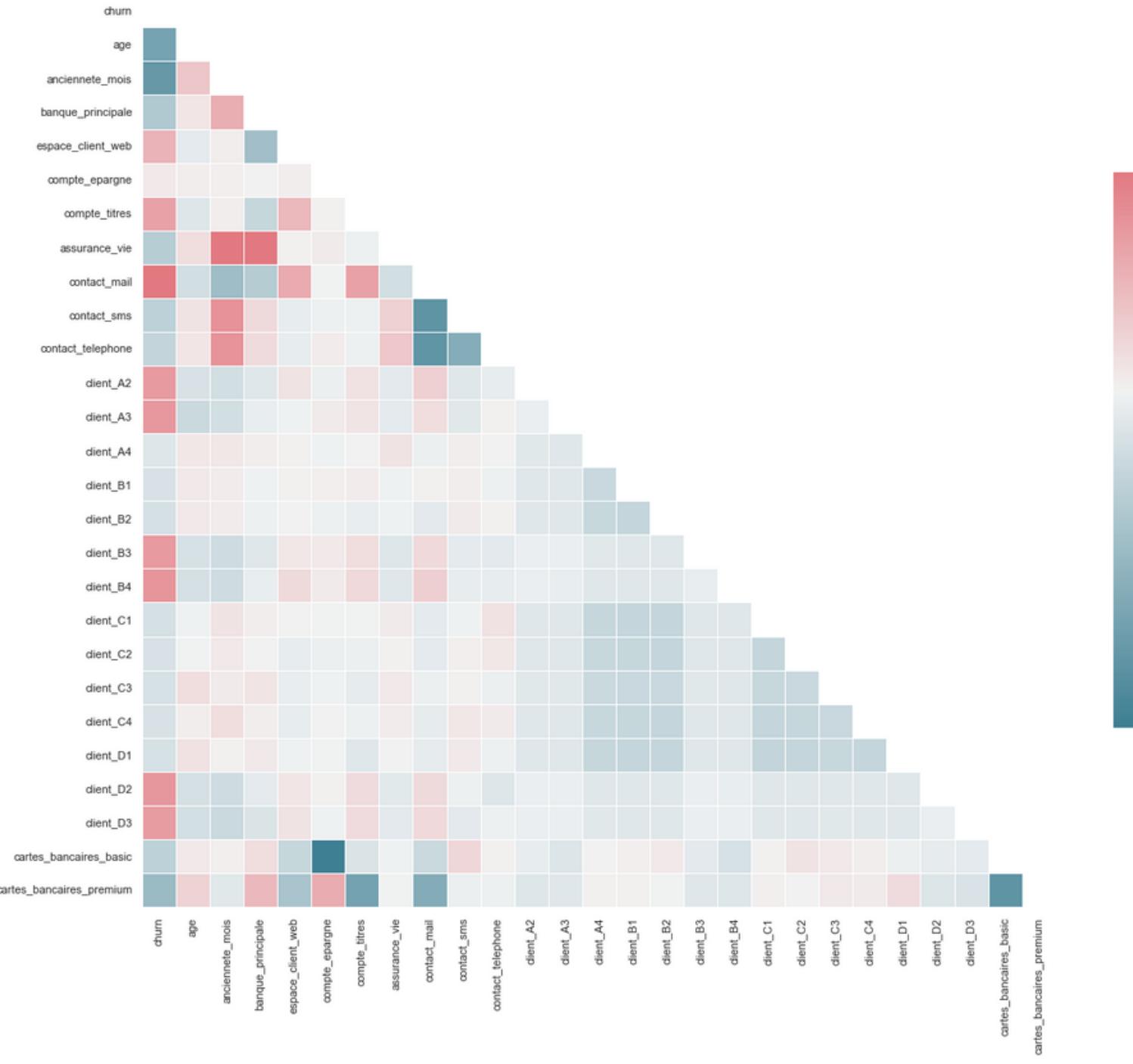
	variables	VIF
0	churn	1.963148
1	age	1.116843
2	anciennete_mois	6.395994
3	agios_6mois	15.949623
4	interet_compte_epargne_total	10.667382
5	banque_principale	1.359054
6	espace_client_web	1.139438
7	compte_epargne	1.819583
8	compte_titres	1.188243
9	assurance_vie	1.440681
10	contact_mail	2.139721
11	contact_sms	1.785195
12	contact_telephone	1.794047
13	client_A2	1.230959
14	client_A3	1.246929
15	client_A4	1.705988
16	client_B1	1.748832
17	client_B2	1.777935
18	client_B3	1.240735
19	client_B4	1.267211
20	client_C1	1.804150
21	client_C2	1.756462
22	client_C3	1.681666
23	client_C4	1.758771
24	client_D1	1.796709
25	client_D2	1.252086
26	client_D3	1.229825
27	cartes_bancaires_basic	20.349631
28	cartes_bancaires_medium	46.358701
29	cartes_bancaires_premium	10.055236

## VIF

- Dans la méthode VIF, nous choisissons chaque caractéristique et la régressions contre toutes les autres.
- Ici la carte medium semble lié aux agios et l'intérêt du compte épargne dépend de l'ancienneté



# Colinéarité entre les variables



Out[63]:

	variables	VIF
0	churn	2.455350
1	age	7.463932
2	anciennete_mois	4.052019
3	banque_principale	1.930413
4	espace_client_web	1.349981
5	compte_epargne	8.564892
6	compte_titres	2.779507
7	assurance_vie	2.799466
8	contact_mail	2.897785
9	contact_sms	2.194765
10	contact_telephone	2.179769
11	client_A2	1.222518
12	client_A3	1.231440
13	client_A4	1.694291
14	client_B1	1.751932
15	client_B2	1.758643
16	client_B3	1.233561
17	client_B4	1.257999
18	client_C1	1.794831
19	client_C2	1.743078
20	client_C3	1.674170
21	client_C4	1.760543
22	client_D1	1.794421
23	client_D2	1.242084
24	client_D3	1.215051
25	cartes_bancaires_basic	1.917663
26	cartes_bancaires_premium	1.902642

VIF

- Nous n'avons plus de variables colinéaires, on peut lancer les modèles pour prédire le churn

# Modèles Machines Learning

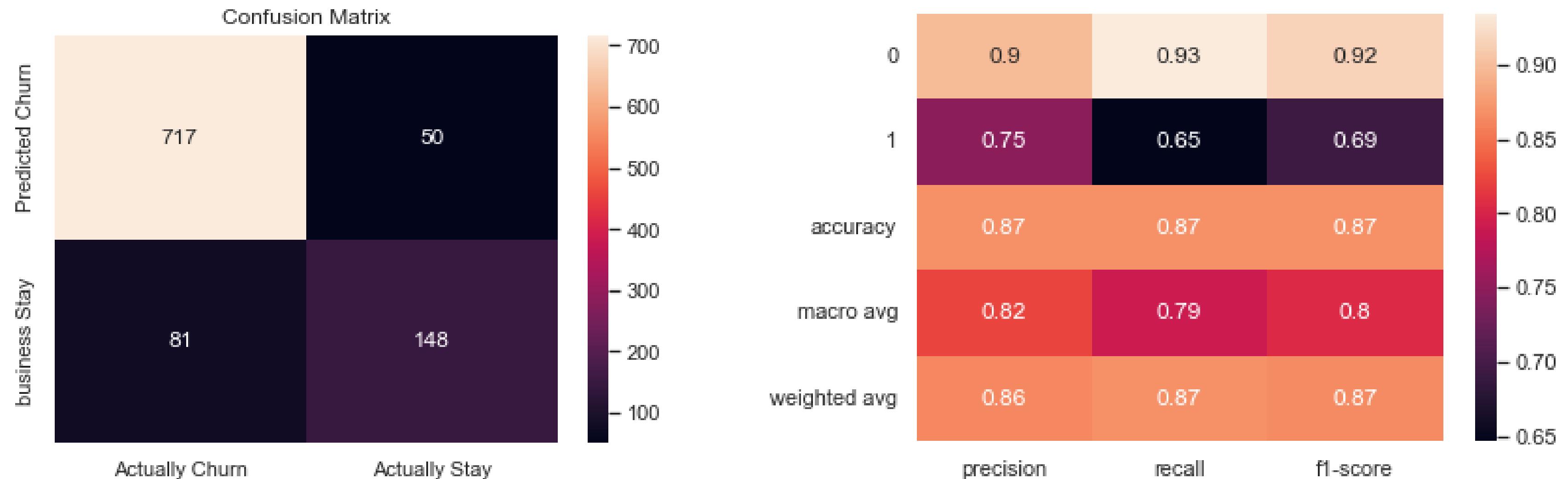
## Random Forest et Logit



- **Précision** : Proportion de clients que l'algorithme prédit partir et qui le font bel et bien.
- **Rappel** : Proportion de clients qui sont partis que l'algorithme à bien identifié comme étant partis.

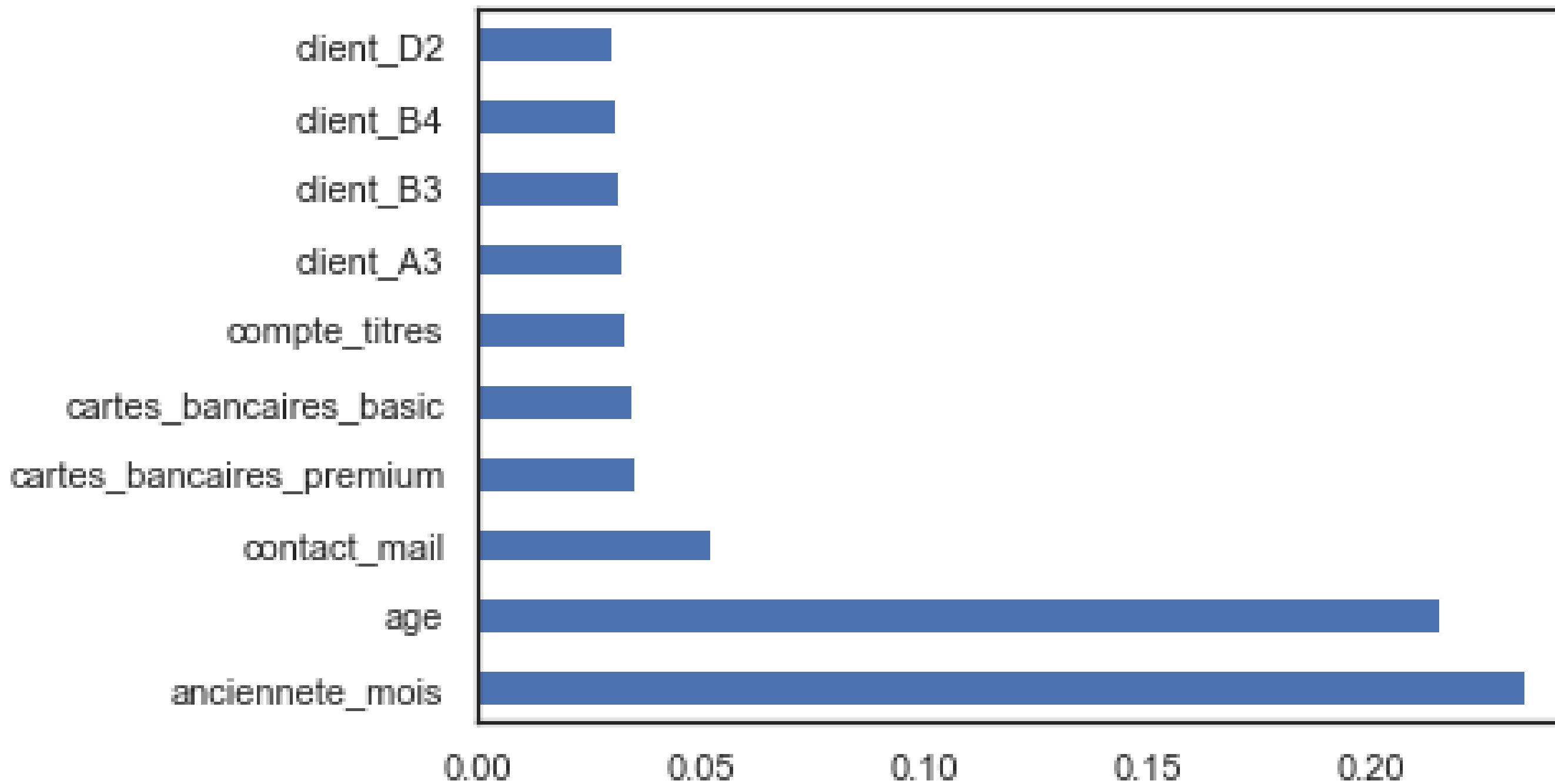


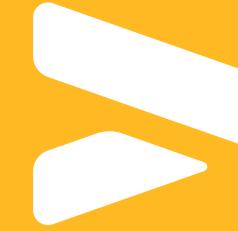
# Random Forest



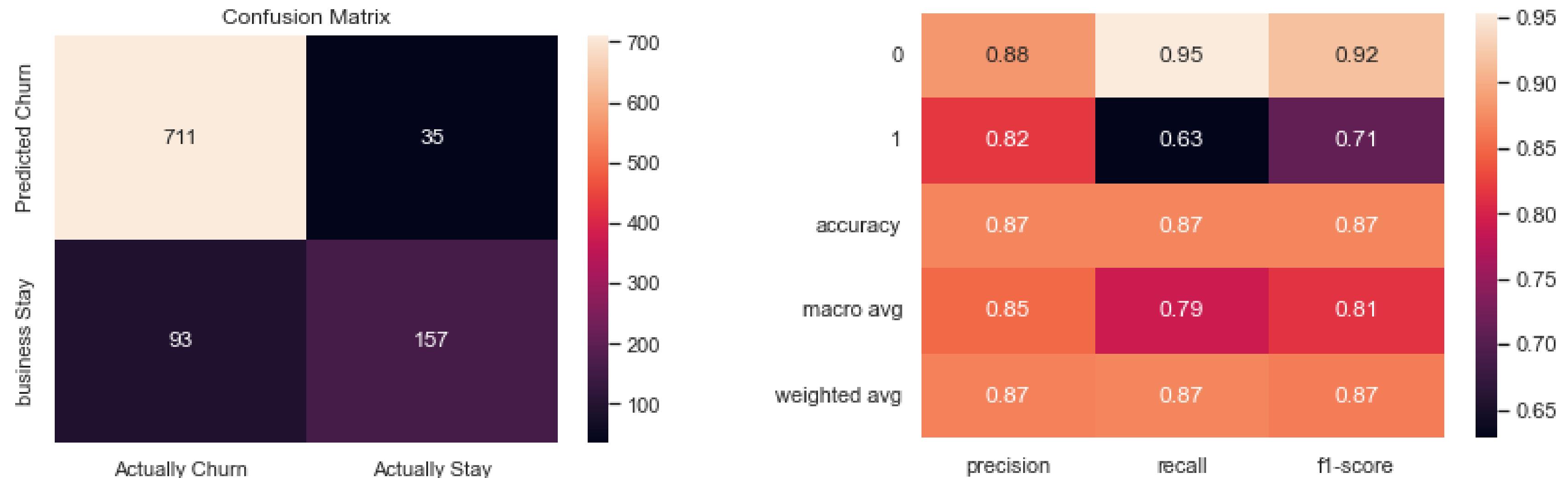


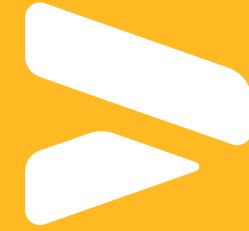
# Most Important Variables





# Logistic Regression





# Score : Probability of churn

Probability\_of\_churn - DataFrame

Index	id client	Probability of churn
0	3463	0.50340647...
1	1869	0.00131046...
2	3760	0.44890970...
3	1463	0.61915230...
4	688	0.02321922...
5	2478	0.33061745...
6	242	0.00659595...
7	3395	0.10747066...
8	3150	0.07453530...
-		

# Merci !

